Assembly of Bacteriophage PRD1 Spike Complex: Role of the Multidomain Protein P5[†]

Javier Caldentey,* Roman Tuma, and Dennis H. Bamford

Institute of Biotechnology and Department of Biosciences, Viikki Biocenter, P.O. Box 56 (Viikinkaari 5), 00014 University of Helsinki, Finland

Received March 29, 2000; Revised Manuscript Received June 12, 2000

ABSTRACT: The spike structure of bacteriophage PRD1 is comprised of proteins P2, P5, and P31. It resembles the corresponding receptor-binding structure of adenoviruses. We show that purified recombinant protein P5 is an elongated (30 × 2.7 nm; R_h = 5.5 nm), multidomain trimer which can slowly associate into nonamers. Cleavage of the 340 amino acid long P5 with collagenase yields 2 fragments. The larger, 205 amino acid long C-terminal fragment appears to contain the residues responsible for the trimerization of the protein, whereas the smaller N-terminal part mediates the interaction of P5 with the pentameric vertex protein P31 (24 × 2.5 nm, R_h = 4.2 nm). In addition, the presence of the N-terminal sequence is required for the formation of the P5 nonamer. The results presented here suggest that P5 and P31 form an elongated adaptor complex at the 5-fold vertexes of the virion which anchors the adsorption protein P2 (21 × 2.5 nm; R_h = 4.1 nm). Our results also suggest that the P5 trimer forms a substantial part of the viral spike shaft that was previously thought to be composed exclusively of protein P2.

Bacteriophage PRD1 is the prototype organism of the Tectiviridae family (*1*). PRD1-like viruses infect a broad range of Gram-negative bacteria harboring a P, N, or W type conjugative plasmid, which encodes the phage receptor complex (*2, 3*). Among the hosts are *Escherichia coli* and *Salmonella typhimurium*. The virion consists of an outer protein shell of about 65 nm diameter enclosing a protein-rich membrane vesicle that closely follows the interior of the icosahedral capsid (*4*). The phage genome, a linear 14 925 bp[1] long double-stranded DNA molecule with inverted terminal repeats and a terminal protein covalently linked to both 5′ ends, is located inside the membrane vesicle (*5*−*8*). Replication of the DNA proceeds via a protein-priming mechanism that includes a sliding-back step (*9*−*11*).

Protein P3 is the major capsid protein and comprises about 70% of the total virion protein mass (*4, 12*). P3 forms a stable trimer, which is the basic building block of the capsid (*13, 14*). Three-dimensional image reconstruction of cryo-electron micrographs has shown that the PRD1 shell contains 240 copies of the P3 trimer arranged on a pseudo T = 25 lattice, an arrangement similar to that of the adenovirus capsid (*4*). Furthermore, the fold of P3 is similar to that of the adenovirus major coat protein, the hexon (*15, 16*). In addition to P3, the outer capsid shell contains a receptor-binding/DNA delivery complex that is composed of proteins P2, P31, and P5, and is located at the 5-fold vertexes (*17*). The observed similarities between PRD1 and adenovirus have led to the surprising conclusion that these viruses share common ancestry (*16*, *18*, *19*).

The morphogenesis of PRD1 is a rather complex process due to the presence of the membrane. During late protein synthesis, soluble capsid proteins appear in the cell cytoplasm, whereas viral membrane-associated proteins are inserted into the host cytoplasmic membrane (*20*). With the assistance of the viral assembly factors P10 and P17 (*20, 21*), coat and membrane proteins interact, giving raise to empty viral particles. On the basis of cryo-electron microscopy and Raman spectroscopic observations (*4*, *14*, *22*), it has been proposed that the viral membrane acts as a scaffold for shell formation. The correct folding of protein P3 requires the *E. coli* chaperonins GroEL/GroES (*23*). In addition, the process of shell formation is coupled to a final step in the folding of the P3 trimer, whereby a stretch of about 20 amino acids forms the interaction between the capsid subunits (*14*).

Our understanding of the assembly of the receptor-binding complex is limited. Recent characterization of the penton protein P31 revealed that its absence led to the lack of proteins P2 and P5, suggesting that these proteins were also located at the 5-fold vertexes (*17*). Protein P2 provides the receptor-binding activity (*24*). Studies performed with small quantities of P5 obtained from dissociated virions indicated that it was a dimer or a trimer (*25*). The sequence of gene *V* also revealed a collagen-like motif (*26*). Additionally, immunoprecipitation experiments have suggested that at least some P5 epitopes are located on the surface of the virion (*27*). Recent work in our laboratory has resulted in the isolation and characterization of mutant phage particles

[†] This work was supported by Grants 168694 and 162993 from the Academy of Finland (Finnish Centre of Excellence Programe 2000-2005).

* Corresponding author. Phone: +358-9-191 59102. Fax: +358-9-191 59079. E-mail: javier.caldentey@helsinki.fi.

[1] Abbreviations: bp, base pair(s); SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis; MALDI-TOF, matrix-assisted laser desorption/ionization time-of-flight; HPLC, high-performance liquid chromatography; UV, ultraviolet; BSA, bovine serum albumin; MDa-kDa-Da, mega-, kilo-, dalton; DLS, dynamic light scattering; CD, circular dichroism; GuHCl; guanidine hydrochloride; ϵ, extinction coefficient; R_h, hydrodynamic radius; M_r, relative molecular mass.

10.1021/bi000711+ CCC: $19.00
Published on Web 08/03/2000

lacking P5 (*28*). These results show that P5 is not involved in P3 shell assembly or DNA packaging but prevents premature release of the DNA. Moreover, its absence had no effect on P31 but resulted in the lack of protein P2.

Successful expression of P5 in *E. coli* allowed us to further characterize its shape, domain structure, and interactions with other vertex proteins. We show that purified P5 is an elongated multidomain trimer, which can associate into higher oligomers. P5 also forms a stable complex with the penton protein P31. The elongated P5 trimer together with P31 forms a substantial part of the virus spike shaft. A model for the assembly of the spike complex is proposed.

MATERIALS AND METHODS

Protein Purification. Protein P5 was overexpressed in *E. coli* HMS174(DE3) cells harboring plasmid pJB51 containing phage PRD1 gene *V*, as described in (*27*). For purification of the protein, induced cells were disrupted by several passages through a French pressure cell (15 000 psi), the cell debris was removed by centrifugation (Sorvall SS-34 rotor, 10 000 rpm, 20 min, 5 °C), and the supernatant was further cleared (Sorvall T865 rotor, 20 000 rpm, 4 h, 5 °C). Proteins were precipitated by addition of cold saturated ammonium sulfate to 25% saturation and subsequent centrifugation (SS-34, 10 000 rpm, 30 min, 5 °C). The pellet was resuspended in 20 mM sodium acetate buffer, pH 5.0, containing 50 mM NaCl, and dialyzed at 5 °C against the same buffer. After clearing (SS-34, 10 000 rpm, 10 min, 5 °C), the protein solution was applied to a Source 15Q column (Pharmacia, Uppsala, Sweden), and proteins were eluted with a linear 50 mM−1 M NaCl gradient. Fractions containing P5 were pooled, and the protein was precipitated by addition of ammonium sulfate as described above. After centrifugation, the pellet was resuspended into a small volume of 20 mM Hepes, pH 7.0, 150 mM NaCl and loaded onto a Hi-Load Superdex 200 16/60 gel filtration column (Pharmacia). Fractions were collected and analyzed by SDS−PAGE. Purified P5 was used immediately or stored in aliquots at −20 °C.

Proteins P2 and P31 were overexpressed in *E. coli* and purified as previously described (*17, 24*).

Mass Spectrometry. MALDI-TOF mass spectrometry was performed with a Biflex mass spectrometer (Bruker-Franzen Analytik, Bremen, Germany), as previously described (*29*). When indicated, samples were cross-linked with glutaraldehyde for 30 min and subsequently dialyzed against water (*29*). For cross-linking, the concentration of purified P5 in the samples was adjusted to approximately 0.4 mg/mL, and the glutaraldehyde concentration was 0.4%. In the case of samples containing P5 fragments (see below), which were more diluted, the glutaraldehyde concentration employed was 0.1%.

Light Scattering. Determinations of molecular weight and hydrodynamic radii characterizations were performed on a Waters 600 HPLC system equipped with a PDI2020/DLS light scattering detector (Precision Detectors Inc.) embedded in the thermostated chamber of a Waters 2440 differential refractometer (*30*). Protein concentrations in the eluted peaks were determined by a standard HPLC UV detector using the absorbance at 280 nm and the extinction coefficients reported below. Simultaneous measurement of refractive index changes and protein concentration allowed for accurate determination of the refractive index increment of all samples (see below). These values were subsequently employed in mass determinations (*31*). The 90° light scattering detector was calibrated using BSA (mass 66 400 Da, $dn/dc = 0.187$ mL/g, $\epsilon_{280} = 0.647$ cm^2/mg) as a standard. Mass distributions as well as hydrodynamic radii were computed using the software provided by the manufacturer [PrecisionAnalyze, (*30*)]. All experiments were performed at 30 °C unless indicated otherwise, and the hydrodynamic parameters were corrected for buffer viscosity and temperature.

The kinetics of $(P5)_3$ self-association were measured on samples eluted from a Superdex-200 10/30 column directly into the thermostated flow cell of the PDI2020 at the desired temperature. Both 90° static scattering as well as DLS were measured as a function of time in 20 s intervals. The static signal was subsequently converted into an apparent mass increase using the known calibration constant of the detector and the concentration measured by the UV detector.

Circular Dichroism. CD spectra were recorded on a Jasco (Tokyo, Japan) J-720 spectropolarimeter using a 2 mm path length cell under the conditions indicated in each case. Each spectrum represents an average of 5 scans corrected for the solvent baseline contribution.

Collagenase Digestion. Digestion of protein P5 was performed with collagenase from *Clostridium histolyticum* (Sigma). The reaction was carried out in 100 mM Tris-HCl buffer, pH 8.5, containing 10 mM $CaCl_2$, with an enzyme:P5 ratio between 1:10 and 1:20. The reaction was analyzed both by SDS−PAGE and by gel filtration. The products of the digestion reaction were purified by chromatography through a Hi-Load Superdex 200 16/60 gel filtration column, as described above. Phage PRD1 samples were also incubated with collagenase under the same conditions used for the digestion of P5. In this case, analysis was done by SDS−PAGE and Western blotting, using a polyclonal anti-P5 antiserum (*27*).

Virus Disruption. Disruption of purified phage PRD1 (*32*) by GuHCl was done essentially as described (*33*). Briefly, PRD1 samples in 20 mM potassium phosphate buffer, pH 7.2, at a protein concentration of 4 mg/mL, were incubated with GuHCl (1.5, 2.0, and 2.5 M, final concentrations) for 30 min at 37 °C. In another set of experiments, virus particles in Tris-HCl buffer, pH 8.5, were disrupted after incubation for 10 min at 70 °C (*34*). After incubation, both GuHCl- and heat-treated samples were spun over a sucrose cushion in an air-driven centrifuge, as previously described (*34*). Prior to centrifugation, samples containing GuHCl were diluted with buffer to reach a final denaturant concentration of 0.75 M. Samples were collected after centrifugation and analyzed immediately by gel filtration chromatography.

General Analytical Methods. SDS−PAGE was performed as described in (*35*). Western blotting was done by transferring the proteins onto a poly(vinylidene difluoride) membrane and visualization with the ECL detection system (Biological Industries, Israel). Protein concentrations were determined from their absorption measured at 280 nm. ϵ_{280} values of 0.613 cm^2/mg for the full-length P5 and 0.452 cm^2/mg for the 141−339 C-terminal domain of P5 were determined using the method of Gill and von Hippel (*36*). Similarly, ϵ_{280} values of 0.973 and 1.20 cm^2/mg were determined for proteins P31 and P2, respectively. These values were subsequently used to determine refractive index increments

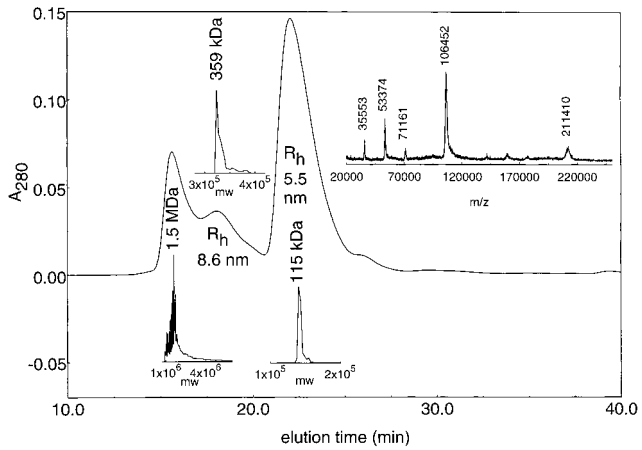

FIGURE 1: Purification and characterization of P5. Elution profile of P5 on a Superdex 200 column measured by absorbance at 280 nm. Insets show mass distributions and apparent hydrodynamic radii obtained by light scattering for proteins in each peak. A MALDI-TOF mass spectrum obtained for the glutaraldehyde cross-linked P5 (peak III) is shown in the upper right corner. Assignments were as follows: *m/z* 35 553, $[M+H]^+$; *m/z* 106 452, $[M_3+H]^+$; *m/z* 53 374, $[M_3+2H]^{2+}$; *m/z* 71 161, $[M_2+H]^+ + [2M+H]^+$; *m/z* 211 410, $[2M_3+H]^+$.

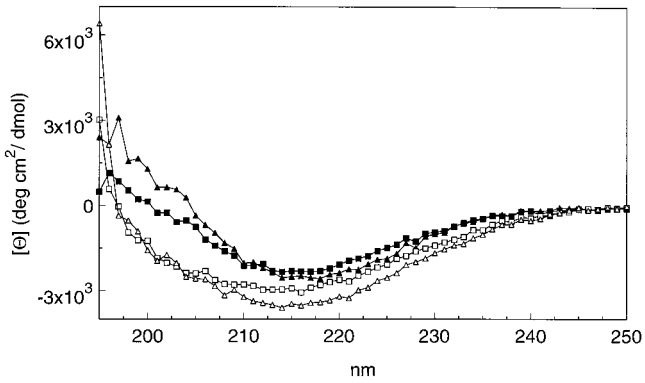

FIGURE 2: Circular dichroism spectra of P5 trimer (squares) and nonamer (triangles), in native (20 °C, filled symbols) and denatured forms (80 °C, empty symbols). Protein concentration was 46 and 53 μg/mL for the trimer and nonamer, respectively.

for all proteins ($dn/dc = 0.179 \pm 0.013$ mL/g for the intact P5, 0.16 ± 0.02 for the C-terminal domain of P5, 0.185 ± 0.012 for P31, and 0.18 for P2). The protein concentration of viral preparations was determined with Coomassie brilliant blue using BSA as a standard (*37*). N-terminal amino acid analysis was done with a PE Applied Biosystems Procise 494A sequencer after SDS−PAGE and blotting (*7*). Analytical gel filtration was performed on a Superdex 200 10/30 column (Pharmacia) that was equilibrated with 20 mM Hepes, pH 7.0, 150 mM NaCl, using a flow rate of 0.5 mL/min. The molecular mass standards used (Pharmacia) were catalase (232 kDa), aldolase (158 kDa), BSA (66 kDa), chymotrypsinogen (25 kDa), and ribonuclease A (14 kDa). Rate zonal centrifugation was carried out in 5−20% (w/v) sucrose made in 20 mM potassium phosphate buffer, pH 7.2, using a Beckman SW 41 rotor (32 000 rpm, 23 h, 10 °C). Ribonuclease A, BSA, and PRD1 P3 trimer [130 kDa, (*15*)] were used as molecular mass markers.

RESULTS

Purification and Characterization of Protein P5. For the purification of protein P5 from induced *E. coli* HMS(DE3)-(pJB51) cells, we followed the procedure described in (*27*) with several modifications in buffer composition and pH described under Materials and Methods. Upon gel filtration chromatography, protein P5 eluted in three distinct peaks (Figure 1), suggesting the presence of various forms of the protein. Determination of elution times in a calibrated column indicated that the first peak (peak I) eluted close to the exclusion volume of the column (over 600 kDa) and thus represents a very large aggregate. The elution times of peaks II and III suggested molecular masses of approximately 300 and 150 kDa, respectively. The protein in all three peaks reacted in Western blots with rabbit polyclonal antibodies raised against the purified P5 (*27*) and yielded the N-terminal amino acid sequence Ala-Asn-Gln-Gln-Ile-Gly-Gly-Ser-Thr-Val, identical to that deduced from the sequence of gene *V* without the first Met residue (*7*).

To further characterize the purified protein, samples from each peak were cross-linked with glutaraldehyde and analyzed by MALDI-TOF mass spectrometry (Figure 1, inset). Non cross-linked samples were used as controls. The experiments carried out with the cross-linked samples indicated that spectra corresponding to peaks I, II, and III displayed major signals at *m/z* 106 606, 106 485, and 106 452, respectively, that were assigned to $[M_3+H]^+$. As shown in Figure 1, weaker *m/z* signals assigned to $[2M_3+H]^+$, $[M_3+2H]^{2+}$, $[M+H]^+$, and $[M_2+H]^+ + [2M+H]^+$ were also detected. In contrast, non-cross-linked samples yielded major *m/z* signals at *m/z* 34 358, 34 376, and 34 210, for peaks I, II, and III, respectively, that were assigned to $[M+H]^+$ (much weaker signals, assigned to $[2M+H]^+$ and $[M+2H]^{2+}$, were also obtained; not shown). These latter values indicate a molecular mass very close to the value of 34 300 Da predicted from the sequence of the protein. Taken together, these results strongly suggest that a trimer constitutes the smallest association unit of P5 and that peaks I and II arise from further oligomerization of $(P5)_3$ multimers.

To confirm the oligomeric state of P5 under native conditions, aliquots of the three peaks were further separated on an analytical Superdex S-200 column directly coupled to a combination of UV, refractive index, 90° Rayleigh, and DLS detectors as described under Materials and Methods. The results of these measurements indicated that the P5 form eluting in peak III had an apparent molecular mass of 115 ± 16 kDa (Figure 1, inset) and a hydrodynamic radius $R_h = 5.5 \pm 0.5$ nm, thus confirming that the trimer is the smallest form of the protein. The hydrodynamic properties of the trimer could be modeled with a 30 nm long, slim (2.7 nm diameter) cylinder (*38*). The form eluting in peak II had a mass of 359 ± 30 kDa (Figure 1, inset) and $R_h = 8.6 \pm 0.6$ nm, suggesting that this peak results from the association of three P5 trimers. Peak I contained a polydisperse mixture of P5 oligomers (aggregates) with masses ranging from 750 kDa to about 3 MDa (Figure 1, inset). Similarly, the R_h values ranged from 11 to 35 nm for this peak. Thus, the peak contains both unresolved higher oligomers as well as excluded polydisperse aggregates of P5.

Secondary Structure of P5. The CD spectra of different P5 samples were obtained and analyzed (Figure 2). The spectra are characterized by the lack of minima at 208 and 222 nm, typical of α-helices, and the presence of a minimum around 217 nm. This is suggestive of a predominantly β-sheet

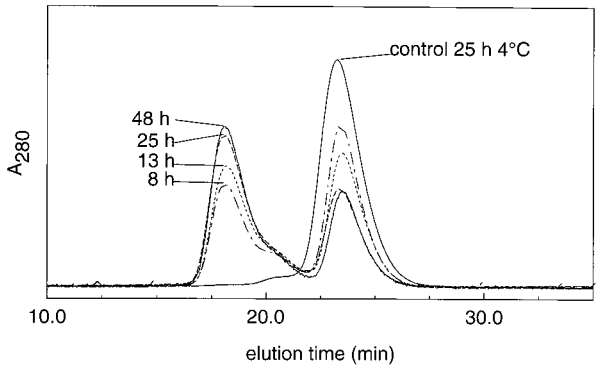

FIGURE 3: Heat-induced oligomerization of the P5 trimer. Shown is the size-exclusion chromatography profile of samples heated for 8−48 h at 37 °C.

secondary structure for protein P5 and is in agreement with secondary structure predictions (>65% β-sheet) based on the amino acid sequence of the protein (*39*). We have compared spectra of the P5 trimer (peak III) and nonamer (peak II), which remain both stable and monodisperse during the duration of the CD experiment. The trimer spectrum displayed a slightly larger ellipticity at 217 nm and a concomitant decrease at 200 nm compared to the nonamer spectrum. This indicates a slightly smaller amount of β-sheet in the trimeric form of P5. When spectra were recorded at different temperatures, no changes could be detected during heating between 20 and 60 °C. Above 60 °C, however, there was a decrease in ellipticity with a concomitant shift of the minimum to approximately 212 nm, indicating aggregation as well as conversion of the native structure into random coil.

Self-Association of P5 Trimers at Elevated Temperature. Incubation of $(P5)_3$ (peak III) samples at 37 °C for a period of time ranging from 8 to 48 h resulted in the formation of nonamers $((P5)_3)_3$ (peak II; up to 80% transformation, Figure 3). On the other hand, samples of $(P5)_3$ that were kept frozen or at 4 °C for weeks remained in the trimeric form (not shown). Neither had a 24 h incubation at 20 °C any appreciable effect on the multimeric state of the protein (data not shown). Cooling the samples back to 4 °C resulted in very slow equilibration (requiring several weeks) between $(P5)_3$ and $(P5)_9$.

To obtain the activation energy and molecular order of the self-association reaction, we have measured the kinetics of polymerization as a function of temperature and P5 concentration. The kinetics of the conversion were measured directly by following the increase in light scattering and hydrodynamic radius (*40*). The initial slopes, corresponding to the pseudo-first-order transformation rates, ranged from 0.014 kDa·min^{-1} at 35 °C to 2.35 kDa·min^{-1} at 50 °C (Figure 4a). An apparent activation energy of 68 ± 4 kcal/mol was obtained (Figure 4b). A log−log plot of the pseudo-first-order rate versus concentration revealed a slope close to unity (Figure 4c). Therefore, the true first-order rate is independent of concentration, and effectively the reaction is zero-order. On the other hand, it is expected that the self-association rate would strongly depend on protein concentration. Thus, the rate-limiting step is not the collision of the protein molecules but a conformational change in each molecule that precedes association.

Interaction of P5 with Vertex Protein P31. Bacteriophage PRD1 has a spike structure containing proteins P31 (vertex

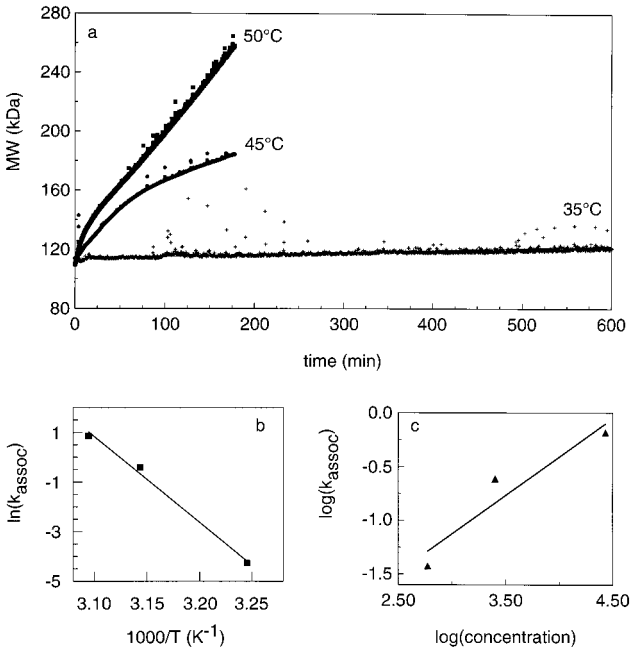

FIGURE 4: Kinetics of the $(P5)_3$ to $(P5)_9$ transformation measured by light scattering at the indicated temperatures. (a) Apparent increase in average molecular mass at 35, 45, and 50 °C for $(P5)_3$ at a concentration of 0.08 mg/mL. (b) An Arrhenius plot of the conversion rate yielded an activation energy of 68 ± 4 kcal/mol. (c) A logarithmic plot of the concentration dependence of rate with a calculated slope of 0.72 ± 0.25.

penton) and P2 (external spike and receptor binding protein) (*17*). Recent results indicate that P5 also forms part of the vertex complex, most probably interacting with P31 (*28*). In the present work, purified proteins P31 and $(P5)_3$ (peak III) were incubated together for different periods of time at 4 °C, and subsequently the samples were analyzed by gel filtration and rate zonal centrifugation. These experiments failed to detect association between the two proteins after incubation for 2 h at 4 °C (Figure 5). However, after a 24 h incubation period, gel filtration chromatography revealed that a fraction of protein P5 in the samples was transformed into a larger, monodisperse complex with an apparent mass of 241 ± 31 kDa and an R_h of 7.8 nm (47 × 3.3 nm; Figure 5a). Analysis of the peak fraction by SDS−PAGE and Coomassie blue staining showed that the complex contained both P5 and P31 (Figure 5a, inset). The apparent mass and the observed decrease in intensity of the P31 peak (mass 68 kDa, R_h = 4.2 nm, modeled as an elongated 24 × 2.5 nm cylinder) suggest that the complex contains six P5 molecules and one to two molecules of P31. This is also consistent with the ratio of gel band densities (Figure 5a). These interactions are specific for P31 since substitution of this protein by BSA or by the PRD1 adsorption protein P2 (mass 66 kDa, R_h = 4.1 nm, elongated 21 × 2.5 nm) did not elicit any change in the association state of P5. Taken together, the results of these experiments suggest that P31 and P5 form a stable and specific complex, in which six molecules of P5 associate with one or two molecules of P31.

When the P5 nonamer (peak II) and P31 were incubated under similar conditions, only a small amount of the $(P5)_6$:P31 complex was formed (Figure 5b). This complex most likely originates from slow dissociation of $(P5)_9$ into $(P5)_3$, which subsequently associates with P31. Alternatively, direct

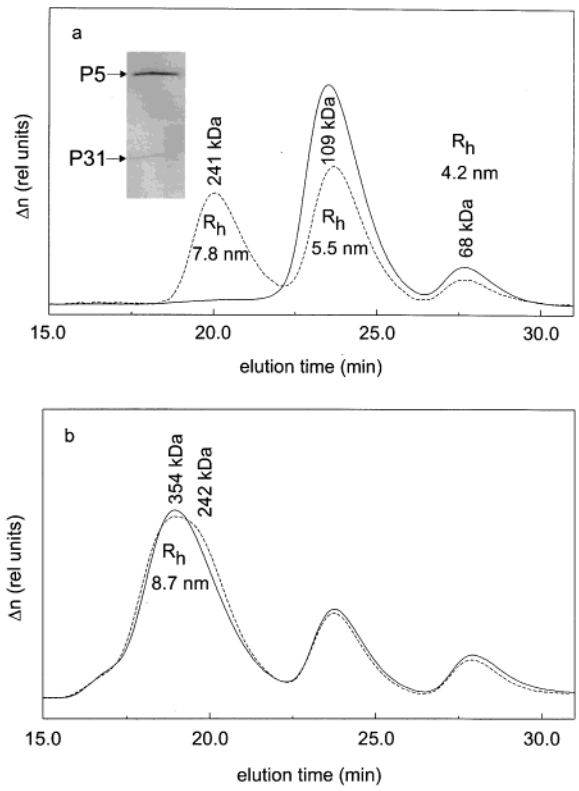

FIGURE 5: Interaction of P5 and P31. (a) $(P5)_3$ was incubated with stoichiometric amounts of P31 (total protein concentration 2 mg/mL) for 2 (solid line) and 48 h (dashed line) at 4 °C, and then separated on a Superdex S-200 column and detected by refractive index (trace shown) and light scattering changes (molecular masses and hydrodynamic radii shown). Peaks with apparent masses of 241, 109, and 68 kDa correspond to the $(P5)_6-(P31)_{1-2}$ complex, $(P5)_3$, and $(P31)_5$, respectively. The inset shows an aliquot of the 241 kDa peak, containing P5 and P31, as analyzed by SDS−PAGE and visualized with Coomassie blue stain. (b) The same as (a) but for $(P5)_9$.

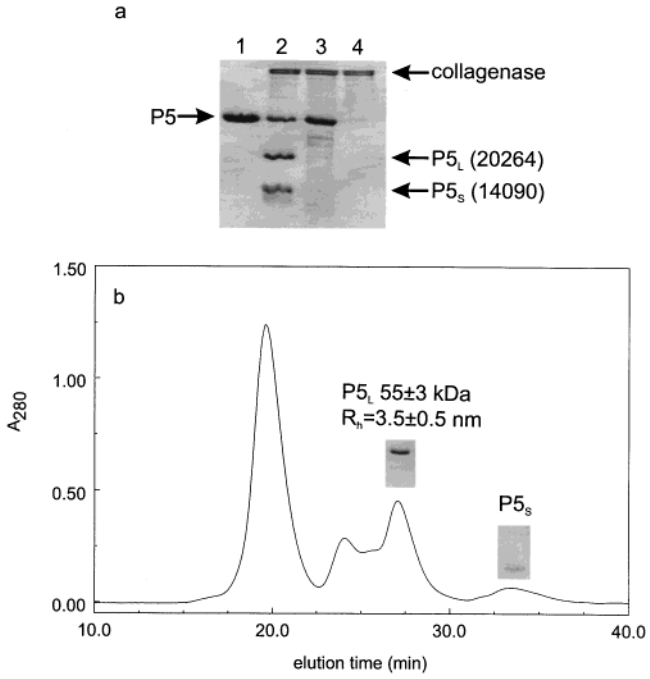

FIGURE 6: (a) Digestion of $(P5)_3$ with collagenase. Lanes: 1, purified P5 without the enzyme; 2, P5 digested overnight at 20 °C; 3, digestion mixture in which P5 had previously been heated to 80 °C; 4, collagenase. (b) Chromatography of the digested products (lane 2 in panel a) on the Superdex S-200 column. The apparent mass and R_h of $P5_L$ were obtained by light scattering.

interaction of the nonamer with P31 dissociates three P5 subunits in order to form the $(P5)_6$:P31 complex.

Oligomeric State of P5 within the Viral Shell. In view of the latter results, we wanted to investigate the state of protein P5 in the viral shell. For this purpose, we disrupted PRD1 particles with GuHCl or by heating. These dissociation procedures, followed by centrifugation at high *g* values, result in the release of shell proteins and their separation from the viral membrane (*33, 34*). When samples of the dissociated proteins were analyzed by gel filtration, protein P5 eluted in all cases at a position corresponding to that of the protein in peak III (not shown). These results suggest that protein P5 is released from the virus as $(P5)_3$. Control experiments performed with purified P5 (peak II) indicated that dissociation of the nonamer does not occur under the conditions employed throughout these tests.

Collagenase Cleaves Protein P5 into Two Fragments. Protein P5 contains a collagen-like sequence, Gly-X-Y, between amino acids 124 and 141 that is susceptible to the collagenase of *Clostridium histolyticum* (*25, 26*). We set out several experiments to further characterize the cleavage of P5 by this enzyme, and the nature of the cleavage products. As shown in Figure 6a, the digestion of P5 yielded two fragments, designated as $P5_L$ and $P5_S$, both of which were recognized on Western blots by the polyclonal anti-P5 antibody (not shown). The larger fragment, $P5_L$, had an M_r of about 20 000, whereas the smaller one, $P5_S$, displayed an M_r of approximately 12 000, as calculated from migration on SDS-containing gels. During the course of these experiments, we noticed that cleavage could be obtained both at 37 °C and at 20 °C. This allowed us to confirm that the three forms of P5 described above were susceptible to the action of the enzyme. Upon incubation of P5 for a period longer than 4 h at 37 °C, further degradation of $P5_S$ into small molecular mass fragments occurred. As shown in Figure 6a, previous heating of protein P5 for 10 min at 80 °C prevented its cleavage by collagenase.

To determine the bond cleaved by the enzyme, we performed N-terminal amino acid analyses of the fragments after their separation by SDS−PAGE and blotting. The sequence Gly-Pro-Ala-Gly-Gly-Thr-Val-Val-Val-Glu was obtained for $P5_L$, whereas the sequence corresponding to $P5_S$ was identical to that of the intact protein. This indicated that collagenase cleaved the bond between Pro and Gly at positions 135 and 136, respectively, in the P5 sequence (Figure 7). Thus, the cleavage occurred between residues 12 and 13 of the 18 amino acid long collagen-like sequence.

Gel filtration chromatography of collagenase-digested samples resulted in the separation and purification of both P5 fragments (Figure 6b). These were further characterized by MALDI-TOF mass spectrometry, performed as for the intact protein. The measurements yielded major signals at *m*/*z* 62 439 and 14 906 for the cross-linked samples of $P5_L$ and $P5_S$, respectively (assigned to the $[M_3+H]^+$ and $[M+H]^+$ species, respectively). The corresponding $[M+H]^+$ signals obtained with the non-cross-linked preparations were 20 264 and 14 090, respectively. These results clearly showed that $P5_L$ maintained a trimeric structure after enzymatic cleavage, whereas $P5_S$ was monomeric. The $P5_L$ fragment was further shown to be trimeric in its native form by light scattering

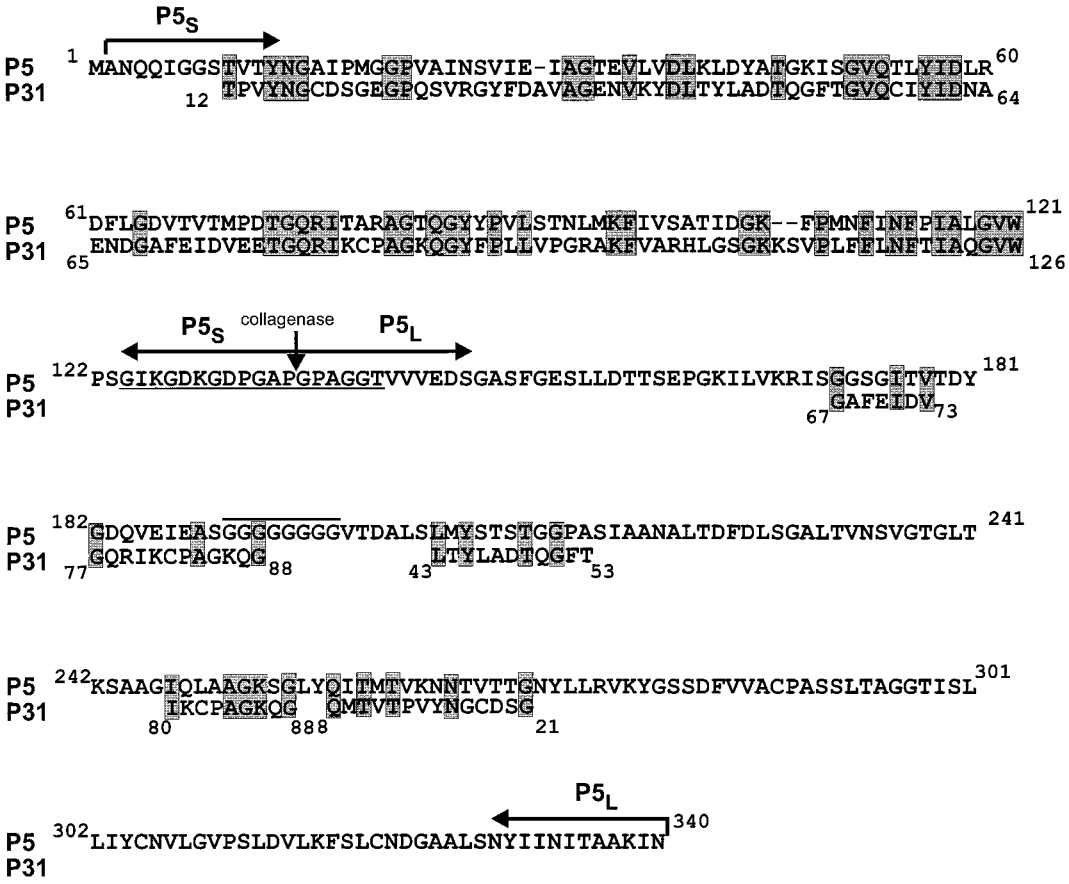

FIGURE 7: Sequence similarity and alignment of proteins P5 and P31. Alignment was done using the ExPaSy server SIM and BLOSUM 62 comparison matrix with a gap penalty of 12 (*43*). Conserved residues are highlighted with boxes. In the N-terminal part of P5 (1−135), the overall identity is 38.3%. The sequences corresponding to $P5_S$ and $P5_L$ are indicated. The collagen-like motif is underlined, and the Gly-rich stretch is overlined. The arrow marks the peptide bond cleaved by collagenase.

(mass 60 kDa, $R_h = 3.5 \pm 0.5$ nm, 13×2.7 nm). No further oligomerization was observed for this fragment at 40 °C (Figure 4a). Nor had P31 any effect on the multimeric state of $P5_L$.

Low yields and further digestion precluded spectroscopic studies on $P5_S$. However, CD measurements done with $P5_L$ indicated that β-sheet was the predominant secondary structure of the fragment, as was observed with intact P5 (Figure 2). Similarly, the large fragment appeared to be stable at temperatures up to 60 °C.

We tested as well the susceptibility of P5 in the intact virion to the action of the *Clostridium* collagenase. No digestion of phage P5 protein could be detected by reaction on Western blots with the P5 antiserum, even after 24 h incubation at 37 °C. Thus, although previous immunoprecipitation experiments indicated that P5 is located on the surface of the virus (*27*), our results suggest that the collagen-like sequence is not accessible to the enzyme.

DISCUSSION

The most striking feature of bacteriophage PRD1 protein P5 is the presence of a collagen-like motif, suggesting an early evolutionary history for the collagen architecture widely used by eukaryotes (*25*, *26*). In the current investigation, we have unambiguously shown that the collagenase of *Clostridium histolyticum* acts on the collagen motif of P5, cleaving the peptide bond between Pro 135 and Gly 136 (Figure 7). This cleavage yields two fragments: the small N-terminal $P5_S$ (14 219 Da) and the larger $P5_L$ (20 116 Da) corresponding to the C-terminal part of the protein. Interestingly, sequence analysis (*17*, *25*) reveals that $P5_S$ shares 38% identity with the vertex protein P31 (Figure 7), whereas only weak similarities were found between P31 and $P5_L$. Heating the protein to 80 °C before digestion precluded the enzymatic cleavage of the protein. Because the collagenase can act on peptides with collagen-like sequences that lack the coiled-coil structure (*41*), it is likely that aggregation of the protein at high temperature (cf. CD) renders the collagen-like sequence inaccessible. It is interesting to note that we did not detect any effects of collagenase on the virion-associated P5. Since protein P5 appears to be located at the surface of the virion (*27*), this indicates that the collagen-like sequence is buried in the viral capsid or that interaction with other viral proteins sterically hinders the protease access.

The smallest form of the obtained recombinant P5 is an elongated trimer with a hydrodynamic radius of 5.5 nm (Figure 8). Using the fragment obtained by collagenase digestion, we have shown that stable trimerization of P5 is mediated by the C-terminal $P5_L$. Thus, the collagen-like motif is not essential for trimerization of the protein. The trimer self-associated into elongated nonamers (hydrodynamic dimensions obtained using the cylinder model: 50×3.6 nm) in vitro. The first-order association reaction is governed by a high activation barrier (68 ± 4 kcal/mol) and is dependent on the presence of the N-terminal fragment. This suggests that the rate-limiting step is a conformational change

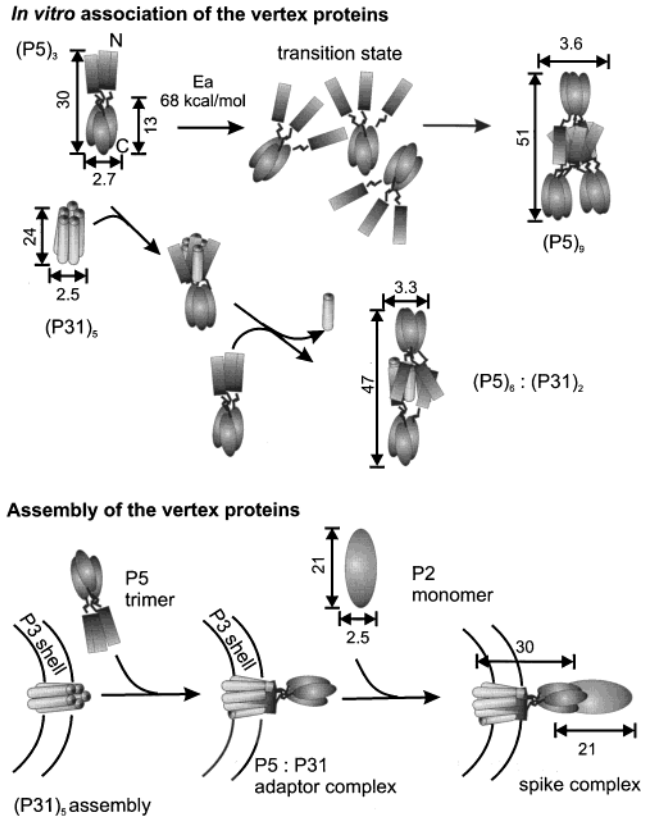

FIGURE 8: Mechanism of P5 transformation and assembly of the PRD1 spike complex. In vitro (upper panel) heating at 37 °C promotes dissociation of the N-terminal domains within the trimer. This constitutes the first-order rate-limiting step with an apparent activation energy of 68 kcal/mol. The "open" N-terminal domain (possibly with assistance of the collagen-like linker) mediates intertrimer association. On the other hand, direct interaction of P31 with the N-terminal domain of P5 facilitates association of two P5 trimers. This interaction is proposed to mediate incorporation of P5 into the spike complex in vivo, as shown in the lower panel. The elongated P5 forms part of the spike shaft, and the C-terminal domain binds protein P2. Hydrodynamic dimensions (nm) are indicated based upon cylindrical models (*38*).

during unfolding of the N-terminal domain, as shown in Figure 8 (upper panel).

In addition, we found that phage PRD1 vertex protein P31 forms an elongated complex with P5 and promotes dimerization of P5 trimers in vitro at low temperature (Figure 5a and Figure 8). Interaction between P5 and P31 has also been detected by the yeast two-hybrid system (M. Aalto, personal communication). The most likely stoichiometry of the complex encompasses six molecules of P5 and one to two molecules of P31 (Figure 8, upper panel). It is conceivable that the observed sequence similarity between P5 and P31 (see Figure 7) is the cause, at least in part, of the interaction between the proteins.

Our data suggest that protein P5 is an elongated trimer composed of at least two domains: the large C-terminal domain responsible for trimerization of the monomer, and the smaller N-terminal domain involved in self-association of the trimeric species and interaction with P31 (Figure 8, upper panel). At this point, it should be noted that the larger $P5_L$ fragment contains a stretch of 8 Gly residues between positions 191 and 198 (*7*) (see Figure 7). Since it has been shown that Gly repeats are typical motifs separating protein domains (*42*), we cannot exclude the possibility that $P5_L$ is itself composed of a larger domain at the most C-terminal portion and a smaller domain containing the collagen-like sequence that confers some additional flexibility at the $P5_S$−$P5_L$ contact region. This added flexibility could be important for the association of $(P5)_3$ with P31 mediated by $P5_S$.

The elongated nature of P5 and P31, as well as $P5_L$, suggests that they are arranged perpendicular to the viral shell and protrude from the rest of the capsid. Furthermore, $P5_L$ resides on the outside of the capsid forming part of the spike shaft (Figure 8, lower panel). This is fully compatible with the accesibility of P5 and P31 epitopes to antigenic recognition within native virions (*17*, *27*). Furthermore, phages deficient in P31 do not incorporate P5 and P2, suggesting that P31 anchors P5 to the viral shell. Recent analysis of P5-deficient phages (*28*) shows that P5 in turn supports the association of the P2 monomer (Figure 8, lower panel). As P2 appears to be also elongated, albeit shorter than the P5 trimer (see Results), the spike shaft is most probably formed by both proteins.

In conclusion, we propose a model for the assembly of the spike complex (depicted in Figure 8, lower panel). Protein P31 interacts with the N-terminal domain of P5. Because the in vitro P5−P31 complex is rather large to be accommodated at the phage vertex, we expect that the P31 pentamer associates with three N-terminal domains to form a larger ring-like structure. The C-terminal domain of P5 forms a trimer and constitutes the spike shaft together with protein P2. In addition, P2 provides the receptor-binding activity. The overall proposed arrangement is reminiscent of that found in the 5-fold vertex of adenovirus, where protein PIII forms the penton base and PIV the trimeric spike shaft and knob. However, in the case of PRD1, the adsorption knob (P2) is most probably monomeric, and the spike shaft is made of two proteins (P5 and P2). This rather complicated vertex assembly may reflect the requirements for DNA ejection from the internal viral membrane of PRD1.

ACKNOWLEDGMENT

Dr. Jari Helin and Dr. Nisse Kalkkinen are thanked for their help with the mass spectrometry measurements and the N-terminal amino acid sequence analyses. Thanks are due to Dr. Sarah Butcher for the gift of some protein P2 and for critically reading the manuscript. The technical assistance of Mrs. Marja-Leena Perälä and Mrs. Riitta Tarkiainen is gratefully acknowledged.

REFERENCES

1. Bamford, D. H., Caldentey, J., and Bamford, J. K. H. (1995) *Adv. Virus Res. 45*, 281−319.
2. Olsen, R. H., Siak, J. S., and Gray, R. H. (1974) *J. Virol. 14*, 689−699.
3. Lyra, C., Savilahti, H., and Bamford, D. H. (1991) *Mol. Gen. Genet. 228*, 65−69.
4. Butcher, S. J., Bamford, D. H., and Fuller, S. D. (1995) *EMBO J. 14*, 6078−6086.
5. Bamford, D., McGraw, T., MacKenzie, G., and Mindich, L. (1983) *J. Virol. 47*, 311−316.
6. Bamford, D. H., and Mindich, L. (1984) *J. Virol. 50*, 309−315.
7. Bamford, J. K. H., Hänninen, A. L., Pakula, T. M., Ojala, P. M., Kalkkinen, N., Frilander, M., and Bamford, D. H. (1991) *Virology 183*, 658−676.

8. Savilahti, H., and Bamford, D. H. (1986) *Gene 49*, 199−205.
9. Savilahti, H., Caldentey, J., Lundström, K., Syväoja, J. E., and Bamford, D. H. (1991) *J. Biol. Chem. 266*, 18737−18744.
10. Caldentey, J., Blanco, L., Savilahti, H., Bamford, D. H., and Salas, M. (1992) *Nucleic Acids Res. 20*, 3971−3976.
11. Caldentey, J., Blanco, L., Bamford, D. H., and Salas, M. (1993) *Nucleic Acids Res. 21*, 3725−3730.
12. Davis, T. N., Muller, E. D., and Cronan, J. E., Jr. (1982) *Virology 120*, 287−306.
13. Luo, C., Hantula, J., Tichelaar, W., and Bamford, D. H. (1993) *Virology 194*, 570−575.
14. Tuma, R., Bamford, J. K. H., Bamford, D. H., Russell, M. P., and Thomas, G. J., Jr. (1996) *J. Mol. Biol. 257*, 87−101.
15. Stewart, P. L., Ghosh, S., Bamford, D. H., and Burnett, R. M. (1993) *J. Mol. Biol. 230*, 349−352.
16. Benson, S. D., Bamford, J. K. H., Bamford, D. H., and Burnett, R. M. (1999) *Cell 98*, 825−833.
17. Rydman, P. S., Caldentey, J., Butcher, S. J., Fuller, S. D., Rutten, T., and Bamford, D. H. (1999) *J. Mol. Biol. 291*, 575−587.
18. Hendrix, R. W. (1999) *Curr. Biol. 9*, 914−917.
19. Belnap, D. M., and Steven, A. C. (2000) *Trends Microbiol. 8*, 91−93.
20. Mindich, L., Bamford, D., McGraw, T., and Mackenzie, G. (1982) *J. Virol. 44*, 1021−1030.
21. Caldentey, J., Hänninen, A. L., Holopainen, J. M., Bamford, J. K. H., Kinnunen, P. K. J., and Bamford, D. H. (1999) *Eur. J. Biochem. 260*, 549−558.
22. Tuma, R., Bamford, J. K. H., Bamford, D. H., and Thomas, G. J., Jr. (1996) *J. Mol. Biol. 257*, 102−115.
23. Hänninen, A. L., Bamford, D. H., and Bamford, J. K. H. (1997) *Virology 227*, 207−210.
24. Grahn, A. M., Caldentey, J., Bamford, J. K. H., and Bamford, D. H. (1999) *J. Bacteriol. 181*, 6689−6696.
25. Bamford, J. K. H., and Bamford, D. H. (1990) *Virology 177*, 445−451.
26. Bamford, D. H., and Bamford, J. K. H. (1990) *Nature 344*, 497.
27. Hänninen, A. L., Bamford, D. H., and Bamford, J. K. H. (1997) *Virology 227*, 198−206.
28. Bamford, J. K. H., and Bamford, D. H. (2000) *J. Virol.* (in press).
29. Helin, J., Caldentey, J., Kalkkinen, N., and Bamford, D. H. (1999) *Rapid Commun. Mass Spectrom. 13*, 185−190.
30. Helfrich, J. P. (1998) *Pharmaceut. Lab.*, 34.
31. Johnson, C. S., and Gabriel, D. A. (1994) *Laser Light Scattering*, Dover Publications, New York.
32. Walin, L., Tuma, R., Thomas, G. J., Jr., and Bamford, D. H. (1994) *Virology 201*, 1−7.
33. Bamford, D., and Mindich, L. (1982) *J. Virol. 44*, 1031−1038.
34. Caldentey, J., Luo, C., and Bamford, D. H. (1993) *Virology 194*, 557−563.
35. Olkkonen, V. M., and Bamford, D. H. (1989) *Virology 171*, 229−238.
36. Gill, S. C., and von Hippel, P. H. (1989) *Anal. Biochem. 182*, 319−326.
37. Bradford, M. M. (1976) *Anal. Biochem. 72*, 248−254.
38. García de la Torre, J., and Bloomfield, V. A. (1981) *Q. Rev. Biophys. 14*, 81−139.
39. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol. 120*, 97−120.
40. Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A., and Teplow, D. B. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 1125−1129.
41. Van Wart, H. E. (1998) in *Handbook of Proteolytic Enzymes* (Barrett, A. J., Rawlings, N. D., and Woessner, J. F., Eds.) CD-Rom, Section 368, Academic Press, New York.
42. Wu, L. F., Tomich, J. M., and Saier, M. H., Jr. (1990) *J. Mol. Biol. 213*, 687−703.
43. Huang, X., and Miller, W. (1991) *Adv. Appl. Mathematics 12*, 337−357.

BI000711+